Biometals
https://doi.org/10.1007/s10534-018-0095-5

Genome-wide and comparative analysis of *bHLH38*, *bHLH39*, *bHLH100* and *bHLH101* genes in *Arabidopsis*, tomato, rice, soybean and maize: insights into iron (Fe) homeostasis

Fırat Kurt · Ertugrul Filiz

Received: 11 January 2018 / Accepted: 13 March 2018

Abstract Iron (Fe) is an essential element for plant life. Its deficiency impedes growth and development and excessive iron can cause the toxic effect via the Fenton reaction. Thus, plants have developed various mechanisms to acquire, distribute and utilize Fe for the maintenance of their iron homeostasis at cellular and systemic levels. A basic helix-loop-helix (*bHLH*) transcription factor family plays essential roles in many regulatory and development processes in plants. In this study, we aimed to understand the roles of *bHLH38*, *bHLH39*, *bHLH100* and *bHLH101* genes for Fe homeostasis in *Arabidopsis*, tomato, rice, soybean and maize species by using bioinformatics approaches. The gene/protein sequence analyses of these genes demonstrated that all bHLH proteins comprised helix-loop-helix DNA binding domain (PF00010) with varied exon numbers between 2 and 13. The phylogenetic analysis did not reveal a clear distinction between monocot and dicot plants. A total of 61 *cis*-elements were found in promotor sequences, including biotic and abiotic stress responsiveness, hormone responsiveness, and tissue specific expressions. The some structural divergences were identified in predicted 3D structures of bHLH proteins with different channels numbers. The co-expression network analysis demonstrated that *bHLH39* and *bHLH101* played more important roles in Fe regulation in *Arabidopsis*. The digital expression analysis showed various expression profiles of *bHLH* genes which were identified in developmental stages, anatomical parts, and perturbations. Particularly, *bHLH39* and *bHLH101* genes were found to be more active genes in Fe homeostasis. As a result, our findings can contribute to understanding of *bHLH38*, *bHLH39*, *bHLH100* and *bHLH101* genes in Fe homeostasis in plants.

Keywords Iron homeostasis · Bioinformatics · Co-expression network · bHLH

Introduction

As a microelement, iron (Fe) is an indispensable element for plants to sustain their metabolic processes such as enzymatic activities, respiration and photosynthesis (Marschner 1995). The deficiency or toxicity

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10534-018-0095-5) contains supplementary material, which is available to authorized users.

F. Kurt
Department of Organic Agriculture, Mus Alparslan University Vocational School, Mus, Turkey

E. Filiz (✉)
Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Cilimli, Duzce, Turkey
e-mail: ertugrulfiliz@gmail.com

Published online: 15 March 2018

of iron impedes plant growth and development resulting in loss of crop yield and quality. Particularly, the iron toxicity gives rise to formation of reactive oxygen species (ROS) as a result of fenton reactions which may have lethal effects for cells (Thomine and Vert 2013). Therefore, to shed light on the metabolic pathway of iron and identification of genes and transcription factors who are responsible for uptake of iron in plants are crucial for plant growth, development and improvement of biofortified crops.

Plants have two distinctive mechanisms to take up iron from soils: Strategy I and II. Strategy I is based on reduction of iron from ferric iron (Fe^{+3}) to ferrous iron (Fe^{+2}) to solubilize iron complexes, used by the plant species, nongraminaceous (non-poaceae or non-grass) family, when soil pH is neutral. On the other hand, the chelation of iron, strategy II, is related to formation of Fe chelates through phytosiderophores to generate soluble Fe complexes by graminaceous plants (Brumbarova et al. 2014; Liang et al. 2017). For nongraminaceous plants to solubilize iron during iron deficient conditions, soil pH in rhizosphere should be reduced by upregulation of *AHA2* proton pump gene. Solubilization of iron is followed by activation of ferric-iron reductase gene which reduces ferric iron to ferrous iron. Once ferrous iron was formed, it was transported into epidermis cells of roots through iron regulated transporter 1 (*IRT1*) (Colangelo and Guennoti 2004; Sivitz et al. 2012; Brumbarova et al. 2014, Vatansever et al. 2015; Liang et al. 2017). Contrary to nongraminaceous plants like *Arabidopsis thaliana*, the iron uptake by graminaceous is dependent on formation of phytosiderophores which chelate ferric iron (Colangelo and Guennoti 2004). After chelation, chelate bound iron was carried into apoplast of root epidermis through *IRT1* (Brumbarova et al. 2014). *IRT1* also reported to be involved in Fe uptake in above-ground tissues of plant (Kim and Guerinot 2007). Regulation of iron uptake genes in *Arabidiopsis* which are responsible for acquisition of iron from soil during iron limited conditions is modulated by *FIT* (FER-like Iron-deficiency-induced Transcription Factor), a bHLH protein (Liang et al. 2017). Additionally, another *bHLH* transcription factor, *POPEYE (PYE),* is positively regulated under iron deficiency and control mobilization of iron between plant organs (Long et al. 2010; Brumbarova et al. 2014). *FIT* was assumed to form heterodimers with bHLH38, bHLH39, bHLH100 and bHLH101 to modulate different subset genes and the expression of *AHA2, FRO2* and *IRT1* genes were thought to be induced by *FIT* under low-iron conditions (Celma et al. 2012). However, increasing evidences suggested that the upregulation of *bHLH038*, *bHLH039, bHLH100, bHLH101*, *PYE*, *NRAMP4, ORG1, PYE, NAS4* were *FIT* independent (Wang et al. 2007; Sivitz et al. 2012).

A basic helix-loop-helix *(bHLH)* was first discovered in animals and later was identified in most of eukaryotic organisms. As a transcription factor, *bHLH* super family proteins have crucial role in many regulatory and development processes in animals and plants. This protein family is approximately 60 amino acids long and has two conserved and distinct domains, basic region and helix-loop-helix. The basic region is composed of 10–15 amino acids while HLH is made up of 40 amino acids (Pires and Dolan 2010; Filiz et al. 2017). The binding of HLH region to DNA is modulated through basic domain whereas formation of dimeric structures such as homo or hetero dimers through interaction of bHLH proteins is regulated by HLH region (Murre et al. 1994). The binding of bHLH proteins to DNA take place with recognition of hexanucleotide sequence defined as E box (Toledo-Ortiz et al. 2003). Palindromic G-box is the most common form of E box where is recognized by many Ib subgroup bHLH proteins (Jones 2004).

The studies of characterization of bHLH proteins in *Arabidopsis* showed that there are as many as 139 or 147 *bHLH* genes (Toledo-Ortiz et al. 2003; Heim et al. 2003). *bHLH* genes regulate different transcription factors and induce upregulation of numerous genes. For example, *bHLH34* and *bHLH104* modulates continuation of Fe deficiency response in *Arabidopsis* (Li et al. 2014). Besides, *bHLH115, bHLH38, bHLH39, bHLH100* and *bHLH101* also play important role in Fe homeostasis. Therefore, in this study, Ib bHLH genes/proteins of *bHLH38, bHLH39, bHLH100* and *bHLH101* were analyzed comparatively at genome-wide scale in *Arabidopsis*, tomato, rice, maize and soybean genomes using bioinformatics tools to have more information of these genes in plant metabolism, particularly in iron homeostasis.

Materials and methods

Retrieving *bHLH* gene sequences

At3g56970 (*bHLH38*), At3g56980 (*bHLH39*), At2g41240 (*bHLH100*), At5g04150 (*bHLH101*) genes were retrieved from UniProtKB database (uniprot.org/) and supplied to Phytozome v12.1.4 for Blastp analysis to identify Ib *bHLH* genes against rice, tomato, soybean and maize (phytozome.jgi.doe.gov/pz/portal.html; Goodstein et al. 2012). Protein domains were queried in Pfam 31.0 database (http://pfam.xfam.org/; Sonnhammer et al. 1997).

Sequence and conserved motif analyses

Coding sequence (CDS), exon and intron organization, peptide sequences, the length of amino acid residues, chromosome numbers of *bHLH* genes were retrieved from Phytozome database (phytozome.jgi.-doe.gov/pz/portal.html; Goodstein et al. 2012). ProtParam server was employed to find physio-chemical features of *bHLH* genes in this study (http://web.expasy.org/protparam/; Gasteiger et al. 2005). CELLO server was used prediction of subcellular localization of *bHLH* genes (http://cello.life.nctu.edu.tw/; Yu et al. 2006). Conserved motifs were analyzed using MEME server with five motifs having 5–60 motifs wide as maximum and minimum motif width respectively (meme-suite.org/tools/meme; Timothy et al. 2009).

Phylogenetic analysis

Alignments of Ib bHLH proteins were made in Bioedit V7.0.5 with Clustal W method (Hall 1999). Phylogenetic tree and distance matrix table of *bHLH* genes were generated in MEGA 7 using maximum likelihood (ML) method, based on James-Taylor-Thornton (JTT) model and 1000 bootstrap, and pair wise method respectively (Kumar et al. 2016).

Promoter sequence and co-expression analyses

Cis-regulatory elements were found by using 1000 bp upstream sites of *bHLH* genes in Phytozome and supplied to PlantCare (http://www.bioinformatics.psb.ugent.be/webtools/plantcare/html/; Lescot et al. 2002). The digital expression analyses of *bHLH* genes were performed by using microarray data from Genevestigator database (https://genevestigator.com/gv/index.jsp; Hruz et al. 2008). ATTED-II plant co-expression database (http://atted.jp) was used to construct co-expression network of *bHLH* genes in *Arabidopsis* (Aoki et al. 2016).

3D structures of bHLH proteins

3D structures of studied proteins were predicted by Phyre^2 server at intensive mode (sbg.bio.ic.ac.uk/-phyre2/; Kelley et al. 2015). 3D structures of bHLH38/39/100/101 proteins were aligned through CLICK server by superimposing protein pairs to have alignment details (mspc.bii.a-star.-edu.sg/minhn/; Nguyen et al. 2011). Quality assessments of 3D structures were made by Vadar server (http://vadar.wishartlab.com, Willard et al. 2003). Molecular cavities of 3D structures of bHLH proteins were computed with Beta-Cavity server using beta-complex, a construct derived from Voronoi diagram of atoms (http://voronoi.hanyang.ac.kr/betacavityweb/about.html; Kim et al. 2015).

Results and discussion

Identification of *bHLH38, bHLH39, bHLH100 and bHLH101* genes

To identifty Ib subgroup *bHLH* genes in rice, tomato, soybean and maize, amino acid sequences of *bHLH38/39/100/101* in *Arabidopsis* were used as reference sequences. A total of 13 genes were found at the end of investigation for rice (one), tomato (three), soybean (four) and maize (one) (Table 1). The shortest open reading frame (ORF) was detected in Glyma.19G132600.1 (354 bp) whereas Solyc10g079660.1.1 had the longest ORF (2094 bp). The polypeptide chains length of *bHLH* genes ranged from 117 to 697 amino acids with 13412.40–78715.51 kDa molecular weights. The maximum number of exons was identified as 13 in Solyc10g079660.1.1 whereas the number of exons of other queried genes ranged from two to four. All genes had the same protein domain structure, helix-loop-helix DNA binding domain (PF00010). Theoretical isoelectric points (*pI*) of genes showed a wide variation ranging from 5.63 to 9.77. Sub-cellular localization of putative

Table 1 The details of *bHLH* genes/proteins in *Arabidopsis*, rice, tomato, soybean and maize

Transcript ID (phytozome)	Species	ORF (bp)	Chr. no	Exon no	Protein length (aa)	Domain family	Mol. wt. (kDa)	*pI*	SL
At3g56970	*Arabidopsis*	762	3	2	253	PF00010	28.714	5.63	Nuclear
At3g56980	*Arabidopsis*	777	3	2	258	PF00010	29.003	6.45	Nuclear
At2g41240	*Arabidopsis*	729	2	2	242	PF00010	27.216	5.90	Nuclear
At5g04150	*Arabidopsis*	723	5	3	240	PF00010	27.716	9.77	Nuclear
LOC_Os01g72370.1	Rice	747	1	3	248	PF00010	27.089	5.78	Nuclear
Solyc10g079650.1.1	Tomato	753	2	3	250	PF00010	28.628	6.40	Nuclear
Solyc10g079680.1.1	Tomato	582	2	3	193	PF00010	22.099	9.77	Nuclear
Solyc10g079660.1.1	Tomato	2094	2	13	697	PF00010	78.715	5.49	Nuclear + PM
Glyma.03G130600.1	Soybean	726	3	3	241	PF00010	27.533	7.10	Nuclear
Glyma.03G130400.1	Soybean	726	3	3	241	PF00010	27.588	6.67	Nuclear
Glyma.19G132500.1	Soybean	723	19	4	240	PF00010	27.397	6.61	Nuclear
Glyma.19G132600.1	Soybean	354	19	3	117	PF00010	13.412	7.97	Nuclear
GRMZM2G057413_T01[a]	Maize	747	3	3	248	PF00010	26.791	5.95	Nuclear

[a]*Zea mays* ENSEMBL 18, ORF: Open Reading Frame, SL: Subcellular Localization, PM: Plasma Membrane

genes was usually predicted in nuclear but Solyc10g079660.1.1 gene may also be in plasma membrane. Niu et al. (2017) stated that *bHLH* genes in *Brachypodium distachyon* (BdbHLH) were orthologous in rice, maize and sorghum and it showed a wide range of variation for *pI*, polypeptide chains length, and molecular weight. Similarly, *bHLH* genes in *Salvia miltiorrhiza* and *Arabidopsis* were reported to have similarity in number and *pI* values of *Salvia miltiorrhiza* were between 4.8 and 9.9 (Zhang et al. 2015b). Collectively, the four Ib subgroup *bHLH* genes had variation in terms of studied properties of proteins and genes indicating functional diversities of *bHLH* genes.

Conserved motifs of Ib bHLH proteins

Based on the results of conserved motif analysis of MEME server, the most conserved five motifs of *bHLH* proteins were shown in Table 2. Motif widths ranged from 21 to 50 amino acids. The motif 1 was detected in all proteins and was related to the HLH domain family. All five motifs were detected in Glyma.03G130600.1, Glyma.03G130400.1 and in Glyma.19G132500.1 whereas Glyma.19G132600.1 had only two motifs, motif 1 and motif 3, respectively. Apart from motif 1, other motifs showed different distribution in *bHLH* proteins (Supplemental Fig. 1). Furthermore, conserved motifs of *bHLH38/39/100/101* proteins in queried species were analyzed by multiple sequence alignment program of Clustal W in Bioedit. As a result, 22 amino acids in bHLH regions were detected as shown in Fig. 1. The conserved residues in this study were lysine (K-Lys), leucine (L-Leu), histidine (H-His), asparagine (N-Asn), alanine (A-Ala), glutamic acid (E-Glu), arginine (R-Arg), serine (S-Ser), proline (P-Pro), threonine (T-Thr), tyrosine (Y-Tyr) and isoleucine (I-Ile). Interestingly motif 1, found all examined genes, was only lack of Thr, instead of having Aspartic acid (Asp-D). Plants were reported to have more conserved Ile-20, Asn-21, Leu-24, Gln (Glutamine-Q)-28, Lys-36, Asp-38, Ile-43, Val (Valine-V)-51 and Leu-54 amino acid sequences in bHLH regions (Niu et al. 2017). However, this suggestion differed from our results in terms of presence of conserved Gln, Asp and Val amino acids. Glu-13, Arg-14, Arg-16 and Leu-27 amino acid residues were reported to be conserved in rice, *B. distachyon* and *Arabidopsis* (Niu et al. 2017). Furthermore, Arg-16, Leu-27, and Leu-61 amino acids were found to have highly conserved in tomato (Sun et al. 2015). Hudson and Hudson (2015) reported that soybean had all but one of bHLH subfamily proteins and its bHLH subfamily proteins had high coherence with *Arabidopsis.* Consequently, they had more similar functionality in metabolic processes. Glu-13 and

Table 2 The *bHLH* proteins with the most conserved five motifs

Motif	Width	Identified site no	E-value	Sequence	Protein domain family[a]
1	29	13 of 13	4.1e−246	VKKLNHNASERDRRKKINDLYSSLRSLLP	Helix-loop-helix DNA-binding domain
2	40	12 of 13	5.2e−137	KAPLSDILQCLENNGLYLLNASSSETFGGRVFYNLHFQVE	Not found
3	50	4 of 13	9.6e−077	MVALFSPPVFSTKGWLLEEDPLSYDVSSEYSFPYQFYSPQTQIELEIERS	Not found
4	21	12 of 13	3.3e−055	SDFVVSTSRLNDCEAVVHISS	Not found
5	21	6 of 13	1.0e−016	YKINCEELSERMLYLYEKCEN	Not found

[a]The Meme motifs were searched in Pfam database to find protein domain families

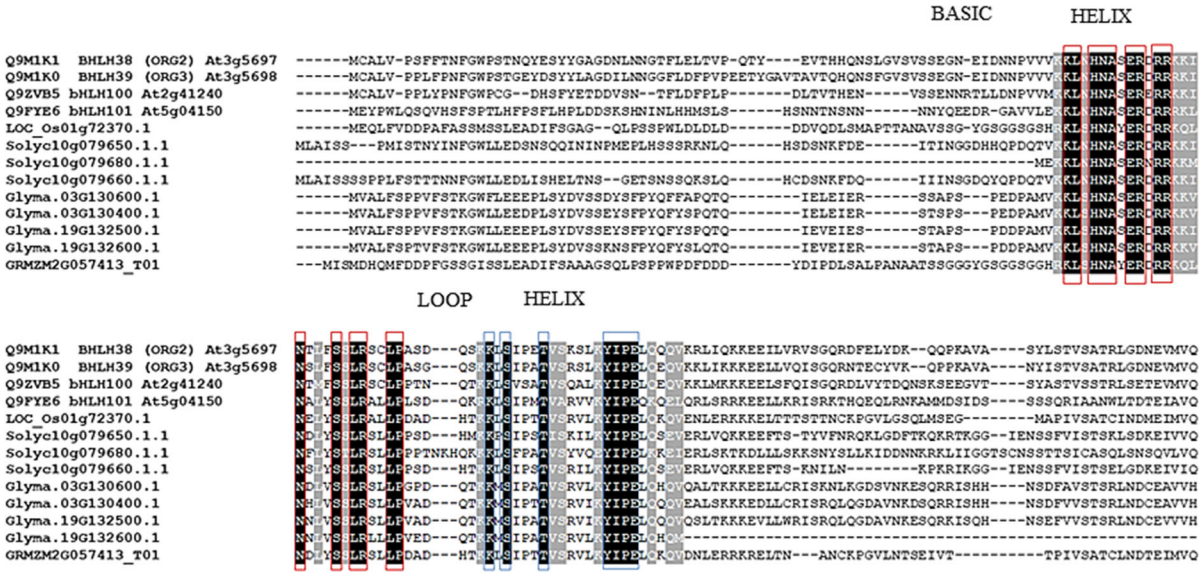

Fig. 1 Multiple sequence alignment of amino acid sequences of bHLH 38/39/100/101 of four species such as *Arabidopsis*, rice, tomato, soybean, and maize. Conserved motifs of sequences were framed in red and blue colors, respectively. The residues of motif 1 were shown in red frame whereas residues of unidentified motif were pointed out with blue rectangular. Identical residues were shown in black columns while similar residues were in gray shading. (At: *A. thaliana*; LOC: *O. sativa,* Solyc: *S. lycopersicum*, Glyma: *G. max*, GRMZM: *Z. mays*)

Arg-16/Arg-17 amino acid residues in the basic region of bHLH were suggested to be involved in DNA binding whereas Leu-27 and Leu-61 residues in helix region was reported to act in dimerization (Sun et al. 2015). Similarly, Pires and Dolan (2010) reported that Ile, Leu and Val amino acids were conserved in most of plant and animal bHLH proteins and stabilization of dimerization were provided by these hydrophobic residues. Also, the first helix was reported to be broken by conserved Pro amino acid. In summary, *bHLH* proteins in this study conserved Ile, Glu, Arg, and Leu amino acids in their HLH regions and these residues might be involved in acquisition of metal ions, particularly iron. Additionally, Met (methionine) and Val were rather found in basic regions of *bHLH* genes but not in all plant species in this study.

Phylogenetic analysis

Phylogenetic tree of bHLH proteins were shown in Fig. 2a. As can be seen, phylogenetic analysis was not revealed a clear distinction between monocot and dicot plants. bHLH proteins were dived into two main groups, group A and B. Group A was composed of two subgroups, A1 and A2. Interestingly, *Arabidiopsis bHLH101* was separated from *Arabidopsis'* other proteins and clustered together with tomato with 85% bootstrap value in subgroup A1. This may be explained by similarity rates between *bHLH* proteins whereby blastp of *bHLH38* were showed 89.5, 73.0 and 67.8% similarity with *bHLH39/100/101* genes in question, respectively. Similarly, Sivitz et al. (2012) reported that bHLH38 and bHLH39 are tandemly situated in chromosome 3 with 79% identity and 89% similarity, respectively. However, *bHLH100* and *bHLH101* are located in different chromosomes and they are 39% identical and 69% similar to each other. To better shed light on this separation we constructed a new phylogenetic tree including BRUTUS (BTS) and PYE genes (Supplemental file, Fig. 2). As a result, bHLH101 was classified into the same cluster with BTS and PYE proteins along with Soly10g079680.1.1. This result showed that *bHLH101* may be outparalogous to PYE, BTS and *Soly10g079680.1.1* genes. Additionally, the internal cluster of tomato was found between Soly10g079650.1.1 and Soly10g079660.1.1 with 91% bootstrap value. The duplication event in tomato *bHLH* genes may be resulted in this paralogous homology. As for Soly10g079680.1.1 and *bHLH101,* they were orthologous to each other. On the other hand, *bHLH100* separated from *bHLH38/39* at 87% bootstrap in A2. Whereas all *bHLH* proteins of dicot soybean were observed in B2, the monocot plants such as LOC_Os01g72370.1 and GRMZM2G057413_T01 were grouped into B1 with 99% bootstrap value.

Gene structure analysis

In terms of exon–intron patterns of *bHLH* genes (Table 1 and Fig. 2b), At5g04150 (*bHLH101*) Solyc10g079650.1.1 and Solyc10g079680.1.1 had two introns while intriguingly Solyc10g079660.1.1 had 12 intron regions. Hudson and Hudson (2015) stated that three intron regions were conserved in majority of *bHLH* genes. Similarly, the *bHLH* intron numbers in tomato were reported to be ranged from null to three and differed from those in *Arabidopsis* in spite of conserved regions of bHLH protein domains (Sun et al. 2015). In this study all *bHLH* genes, except At5g04150, in *Arabidopsis* had one intron. GRMZM2G057413_T0 along with all soybean genes had two introns whereas three introns were observed in LOC_Os01g72370.1. Soybean intron structure in bHLH subgroups were reported to be coherent with those in rice and *Arabidopsis*. In this respect, bHLH family showed null to three conserved introns showing nine patterns depending on intron positions in soybean (Hudson and Hudson 2015). Also, the relationship of *bHLH* genes in *Arabidopsis*, rice, tomato, maize and soybean was complicated. The intron regions of *bHLH* genes in this study showed coherence with the results

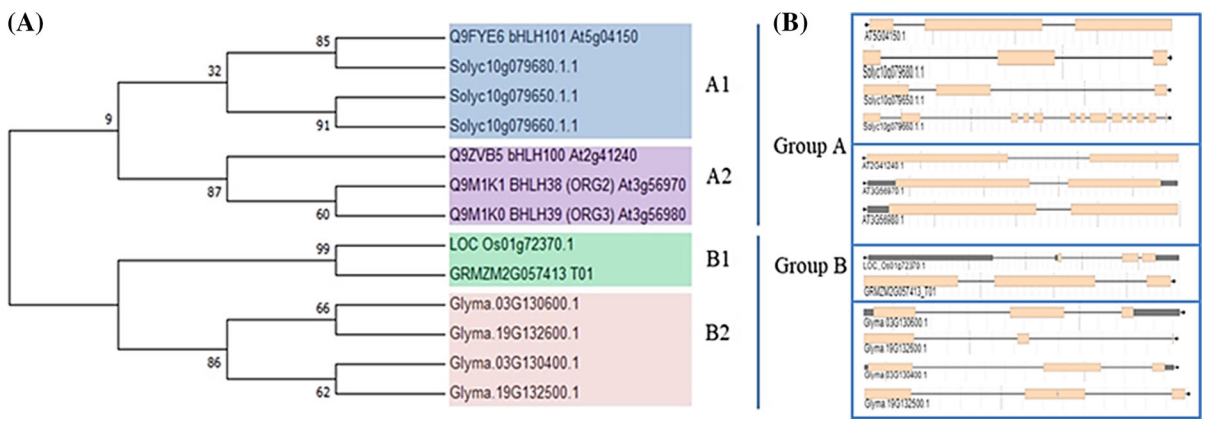

Fig. 2 Phylogenetic tree (**a**) and exon–intron distribution (**b**) of *bHLH38/39/100/101* genes/proteins of *Arabidopsis*, tomato, rice, maize and soybean. The tree was generated in MEGA 7.0 software using maximum likelihood (ML) method with 1000 bootstrap replicates. The yellow filled bars and black lines show the exon and intron structures, respectively in Fig. 2B

of other studies of *bHLH* genes, except Solyc10g079660.1.1. In addition, the distinction of *bHLH* genes in monocot and dicot were not found in phylogenetic tree. According to gene structures and phylogenetic relationships of studied *bHLH* genes, it can be suggested that relationships of *bHLH* genes in *Arabidopsis*, rice, tomato, maize and soybean was complicated.

Promoter region analysis

Cis-acting elements are located in the upstream of gene coding regions within in gene promoters and the regulation of transcription is related to finding motifs in promoter region (Garcia and Finer 2014). To identify *cis*-regulatory elements related to *bHLH* genes, PlantCare database was supplied with 1000 bp upstream DNA sequences of each *bHLH* genes in this study. As a result, a total of 61 *cis*-elements were identified and a heatmap were constructed to give a better insight into *cis*-regulatory elements in associated with *bHLH* genes. *Cis*-regulatory elements whose functions are unknown were excluded from the Heatmap (Fig. 3) and absent or present elements were shown in red and green colors, respectively. The highest number of *cis*- elements was identified in *GRMZM2G057413_T01* and *Glyma.03G130600.1* genes. In this sense, CAAT and TATA boxes were found in all *bHLH* genes of each species investigated. Skn-1_motif was present, except *LOC_Os01g72370.1*, in all genes under investigation. TATA box is part of core promoter region of protein coding genes and is a binding site for TATA binding protein (TBP) which make up transcription initiation factor (TFIID) (Garcia and Finer 2014). CAAT box is a binding site for transcription complex with which modulation of transcription of genes occur (Laloum et al. 2013). Skn-1 motif is required for endosperm expression and has an important role with other *cis*-acting elements such as 5UTR-Py-rich stretch, ABRE, Box-4, G-Box, MBS in the upregulation of genes acting in in oxidative defense pathway in rice (Yousefi et al. 2012). G-Box and G-box were found in more than half of *bHLH* genes in the study, showing agreement with the report that G-box (CACGTG) is a common motif for bHLH and bZIP protein families (Wong et al. 2017). G box is also identified as a MYC recognition site along with MYB, having role in upregulation rd22 drought-induced gene. Moreover, bHLH associated protein AtMYC2 and MYB connected protein AtMYB2 together binds to MYB and MYC binding sites to upregulate ABA inducible genes during drought stress. ABRE sequence was reported to be a promoter region for ABA inducible genes (Abe et al. 2003). In parallel to these reports, MBS and ABRE elements also were found as *cis*-elements in this study, involving in MYB binding site involved in drought-inducibility and ABA responsiveness respectively. Apart from these findings, the other promoter sequences, having role in hormone metabolism for *bHLH* genes, were found as TCA-element (salicylic acid responsiveness), CGTCA-motif (the MeJA-responsiveness), TGA-element (auxin-responsive element), TATC-box (gibberellin-responsiveness), CE3 (*cis*-acting element involved in ABA and VP1 responsiveness), TGACG-motif (involved in the MeJA-responsiveness), GARE-motif (gibberellin-responsive element) and ERE (ethylene-responsive element). Also, a considerable number of *cis*-regulatory elements for *bHLH* genes in the study involved in light responsiveness (Box I, Box 4, AE-box, GA-motif, TCT-motif, GT1-motif,I-box,L-box, Sp1, TCCC-motif, TCCC-motif, as-2-box, ACE, chs-CMA1a, MNF1, rbcS-CMA7a, ATCT-motif, CATT-motif, MRE, AAAC-motif, GATA-motif, chs-Unit 1 m1, Box II, 3-AF1 binding site, AT1-motif, C-box, GTGGC-motif). The other *cis*-regulatory elements of *bHLH* genes can be grouped into three classes: tissue-specific, stress responsive elements and other elements. *Cis*-regulatory elements, revealed in the abiotic stress responsiveness, were ARE (anaerobic induction), LTR (low-temperature responsiveness), TC-rich repeats (defense and stress responsiveness), GC-motif (enhancer-like element involved in anoxic specific inducibility) and HSE (heat stress responsiveness). As for tissue-specific *cis*-regulatory elements, they were identified as CCGTCC-box (related to meristem specific activation), RY-element (seed-specific regulation), CAT-box (meristem expression), as1 (root-specific expression), AACA_motif (endosperm-specific negative expression) and GCN4_motif (endosperm expression). Lastly, several *cis*-regulatory elements of *bHLH* genes in this study were identified functioning in fungal elicitor responsiveness (Box W1), regulation of circadian control (circadian) and regulation of zein metabolism (O2-site). Overall, the diverse functions of *cis*-regulatory elements of *bHLH* genes in various metabolic pathways have shown a complex system

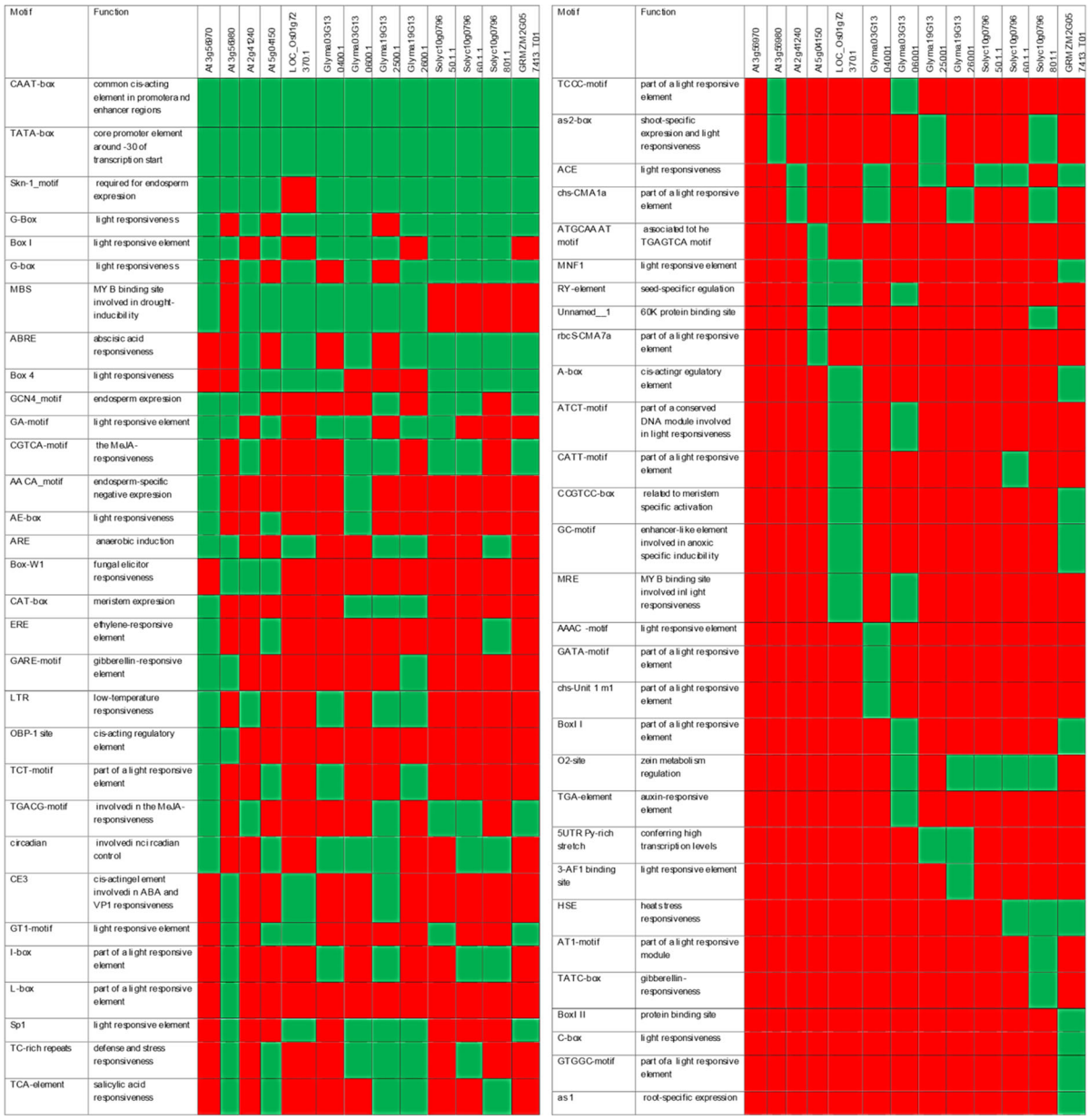

Fig. 3 The heatmap of *cis*-elements in *bHLH38/39/100/101* genes from *Arabidopsis*, rice, soybean, tomato and maize, respectively (green: present and red: absent)

which regulates numerous metabolic processes in investigated plants.

Co-expression network analyses of *bHLH* genes in *Arabidopsis*

The four *A. thaliana* genes such as *bHLH38* (At3g56970), *bHLH39* (At3g56980), *bHLH100* (At2g41240) and *bHLH101* (At5g04150) were investigated in ATTED-II co-expression database to shed a light on interactions among them and on their roles as transcription factors in synthesis of other proteins which modulates numerous biological processes. For this purpose, an interactome map was constructed (Fig. 4). At3g56980 and At5g04150 genes were found in co-expressed gene network whereas no co-

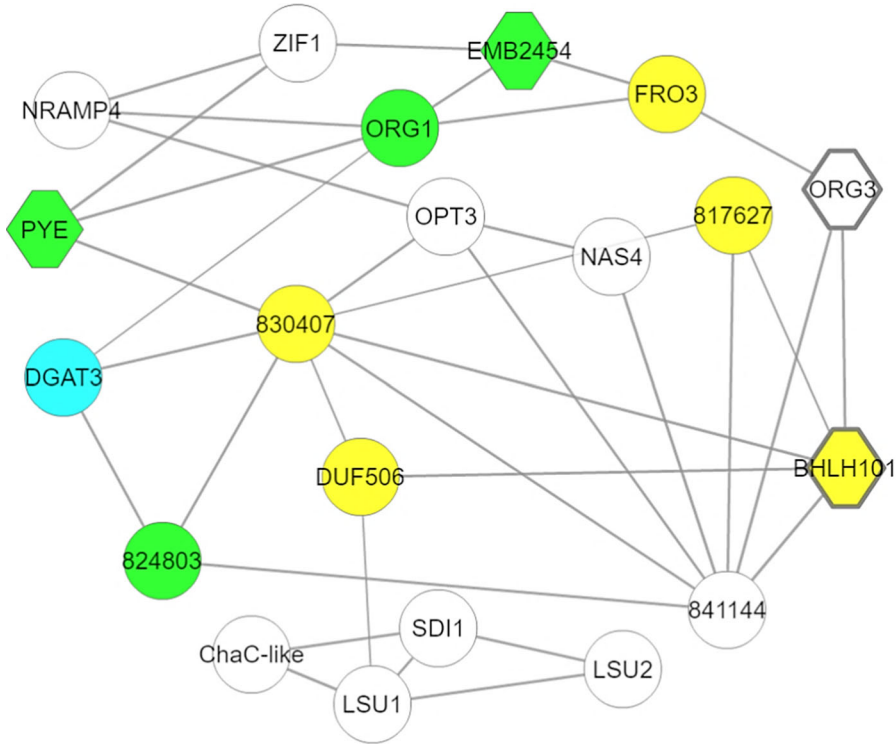

Fig. 4 The co-expression network of *bHLH101* and *bHLH38* (*ORG3*) genes in *Arabidopsis* by ATTED-II server

expression data was found for At3g56970 and At2g41240. At3g56980 (*ORG3 or OBP3-responsive gene 3*), At5g04150, At3g56970 (*ORG2* or *OBP3 responsive gene 3*) and At2g41240 was reported to have functions in the regulation of transcription iron ion homeostasis and cellular response to iron ion starvation (Xiang 2015). At2g41240 was suggested to be expressed in response to drought stress (Rasheed et al. 2016). In co-expression network, At3g56980 (*ORG3*) putatively interacts with At5g04150 (*bHLH101*, another transcription factor), At1g23020 (*FRO3*) and At1g47400 (841144) gene, which has currently not been annotated. At1g23020 (*FRO3*) is defined as iron ion transport, protein involving in oxidation–reduction process and cellular response to iron ion starvation (Xiang 2015). Furthermore, this gene was reported to encode superoxide-producing enzyme NADPH oxidase (Rizhsky et al. 2003). NADPH oxidase, also known as respiratory burst oxidase homologs (Rbohs), named *AtRohA* to *J*, has 10 members in *Arabidopsis*, each of them have different roles and expression patterns. For example, *AtRbohD* and *AtRbohF* were stated to regulate numerous abiotic stresses and pathogen defense response (He et al. 2017). As for At5g04150 (*bHLH101*), it supposedly interacts with At3g56980 transcription factor and At1g12030 (*DUF 506*), At5g05250 (830407), At2g30760 (817627) and At1g47400 (841144) genes whose annotations are not available.

There are two networks regulating iron uptake mechanisms: *POPEYE (PYE)* network and *FIT* network. *PYE* is a bHLH protein involves in redistribution of iron ions in plants (Brumbarova et al. 2014). It downregulates *NICOTIANAMINE SYNTHASE 4* (*NAS4*), *FRO3*, and *ZINC-INDUCED FACILITATOR1* (*ZIF1*) (Liang et al. 2017) and interacts *bHLH105* (*ILR3*) and bHLH115. Meanwhile, the dimers of *bHLH104* and *bHLH105* or *bHLH34* and *bHLH105* involve in modulation of PYE (Conorton et al. 2017). *bHLH105*, *bHLH104* and *bHLH34* act in the regulation of *bHLH38*, *bHLH39*, *bHLH100* and *bHLH101* (Liang et al. 2017). *FIT* is orthologue of *FER*, a *bHLH* transcription factor in tomato with At2g28160 gene number, known also as *FIT1* or *FRU FIT*, expressed in roots under low iron conditions and was suggested as the key regulator of iron mechanism (Wang et al. 2007). Upregulation of *FRO2* or *IRT1* genes were reported not to be dependent on constitutively upregulation *FIT* but homo or hetero-dimerization of *FIT* with *bHLH038, bHLH039, bHLH100,*

bHLH101 genes, whose expression hindered by excessive iron medium (Xing et al. 2015; Zhang et al. 2015a). Transcription of these genes were reported to occur *FIT*-independent way (Wang et al. 2007; Sivitz et al. 2012; Xing et al. 2015). Furthermore, although *bHLH38/39/100/101* were described as functionally redundant genes in *Arabidopsis*, based on results obtained from knockout experiments, the functional redundancy of *bHLH101* with *bHLH38* was not found in a melon mutant, named fefe, under adequate iron conditions (Ramamurthy and Waters, 2017). Similarly, another knockout experiment in *Arabidopsis* showed that presence of *bHLH38* and *bHLH101* or *bHLH38* and *bHLH100* genes did not affect iron uptake mechanism under iron deficiency (Wang et al. 2013). Overall, though *bHLH38, bHLH39, bHLH100* and *bHLH101* genes play important roles in iron uptake mechanism, *bHLH39* and *bHLH101* may be suggested as more crucial genes regulating metal homeostasis, particularly iron uptake and regulation metabolism in plants.

3D structures of bHLH proteins

Understanding or prediction of 3D structure of a protein is crucial to know how it functions in any processes in metabolic pathways. Comprehension the structures of *bHLH38/39/100/101* proteins will give insights into their features such as conformational flexibilities, allosteric regulations, catalytic activities and posttranslational modifications (Eisenhaber 2006). In this respect, 3D structures of bHLH38/39/100/101 proteins were constructed by Phyre^2 server (Fig. 5). Firstly, the seconder structure analyses and quality assessment of protein models were made by Vadar server. The seconder structure analyses revealed that *bHLH* proteins varied 22–42% for α-helices, 0–16% for β-strand, 48–61% for coils, and 4–22% for turns. Furthermore, Ramachandran analysis showed that 81 and 94% of residues were in allowed region, indicating that the quality of models was good. The 3D models were superimposed as pairs to reveal their similarities and divergences based on overlap values. The calculations were made superimposition of each *bHLH* (*bHLH 38/39/100/101*) proteins of *Arabidopsis* against bHLHs in other plant species in this study. Also, to shed light on phylogenetic relationship of *bHLH* proteins, the *bHLH* proteins under the same internal clad in phylogenetic tree was superimposed to each other (refer to phylogenetic section). Thus relying on overlap values, showing similarities and divergences of 3D structures, how functions and structures of these proteins during evolutionary process were changed was estimated (Supplemental file Table 2). In this respect, 49.57, 42.98 and 39.38% overlap values were obtained by superimposition of At2g41240 (bHLH100) with Glyma.19G132600.1, Solyc10g079650.1.1 and Solyc10g079680.1.1 respectively. The superimposion of At3g56970 (bHLH38) with Glyma.19G132600.1, At5g04150 (bHLH101) and Glyma.19G132500.1 showed 52.99, 38.75 and 37.92% similarity respectively. The 47.86, 45.60 and 36.10% overlap values were obtained when At3g56980 (bHLH39) superimposed Glyma.19G132600.1, Solyc10g079680.1.1 and Glyma.03G130600.1 respectively. At5g04150 (bHLH101) showed highest similarities with Glyma.19G132600.1 (51.28%), Glyma.19G132500.1 (40.83%) and At3g56970 (bHLH38) (%39.17). These results showed that *Arabidopsis* bHLH38/39/100/101 were structurally most similar to Glyma.19G132600.1. However, analysis of phylogenetic tree and proteins' structures were not coherent. Though these overlap values found low, the overlap values of *Arabidopsis* genes to other *bHLH* genes in this study were still above the twilight zone (> 30%). In the meantime, to present 3D molecular structures of studied proteins better the molecular cavities' functions were computed (Fig. 5). Properties of cavities (or voids) along with channel numbers also were reported to be important parameters to determine the interactions of polypeptide with other molecules in their environment (Kim et al. 2015). According to channel numbers, bHLH proteins showed a considerable variation ranged from 3 to 11 in all species investigated (Supplemental file Table 3). All data considered, 3D structural variation may be explained by functional diversities of *bHLH* genes in plant metabolic processes.

Analyses of *bHLH38/39/100/101* gene expressions in *Arabidopsis*

The co-regulation of *bHLH* genes was examined in terms of developmental stages, perturbations and anatomical parts. In this respect, microarray data of each *bHLH* genes were retrieved from Genevestigator platform. Based on developmental stage gene

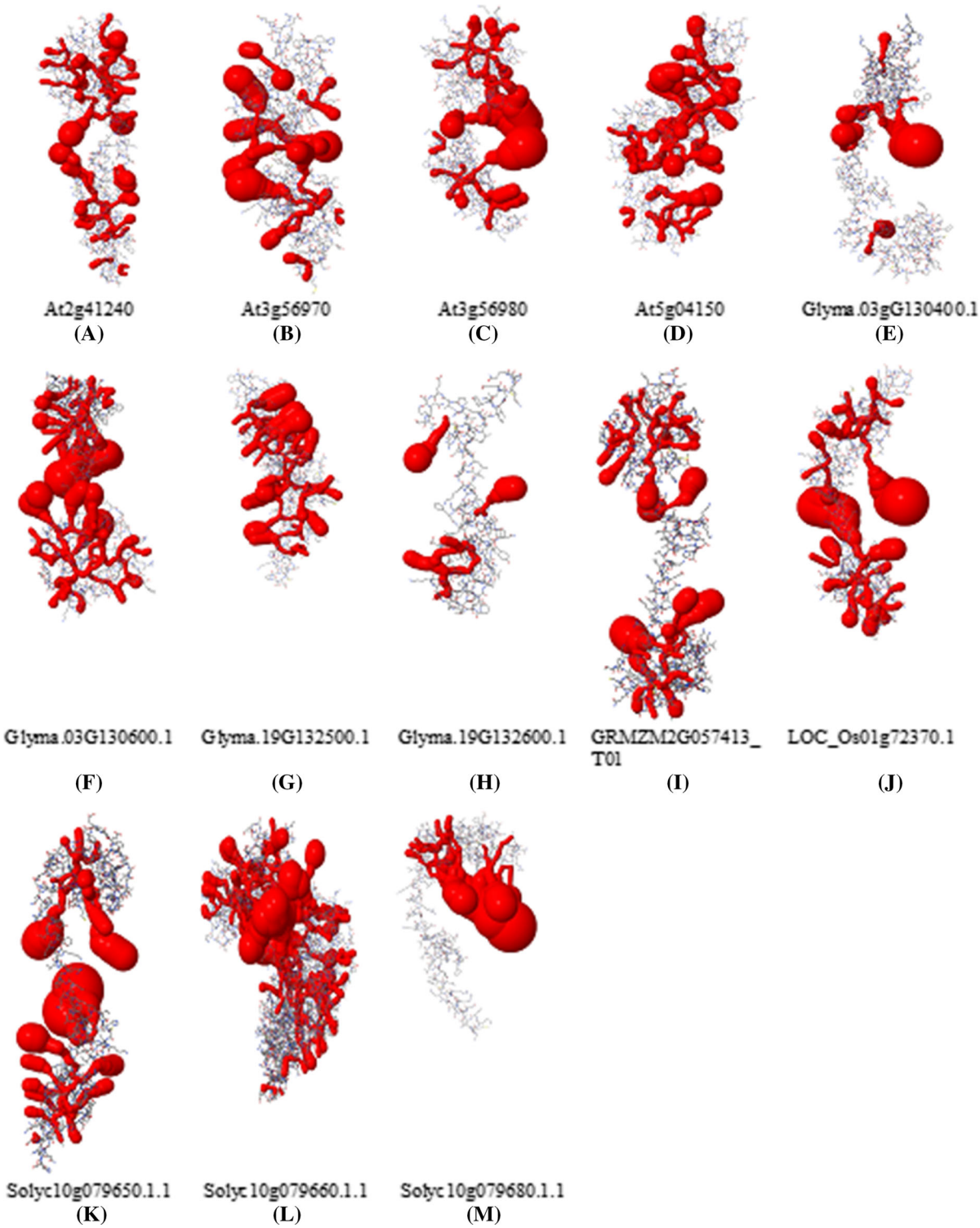

Fig. 5 The predicted 3D structures of bHLH proteins by Phyre[2] server. The putative channels were predicted by Beta-Cavity server and shown as red on 3D structures

expression data, *bHLH38* highly expressed in flowering stage whereas *bHLH39* was expressed at moderate level from seedling through flowers and siliques stage. The main increase of *bHLH39* expression was observed in transition from germination to seedling stage. *bHLH100* was highly expressed in flowering stage while *bHLH101* expression was fluctuated between low and high levels from seedling through flowering stages but showed decrease, after flowering, through senescence stage. Interestingly, although expression levels differed, *bHLH38* and *bHLH100* had same similar expression patterns with regard to expression tendencies as was the case for *bHLH39* and *bHLH101* (Fig. 6). This data made us infer that iron should be supplied to *Arabidopsis* particularly in seedling and flowering stages since upregulation of four Ib subgroup *bHLH* genes were dependent on iron limiting stress conditions (Xing et al. 2015; Zhang et al. 2015a).

Gene expression data of *bHLH* genes in different anatomical parts were presented in Fig. 7. Both *bHLH38* and *bHLH100* genes were expressed high in adult leaves. On the other hand, *bHLH39* and *bHLH101* genes showed more complicated expression profiles in more anatomical parts. For example, *bHLH39* gene was highly expressed in root cell and shoot stele cell along with their lower components such as radicle and hypocotyl, suggesting that iron's involvement in photosynthesis together with maintenance of chloroplast structure and function (Rout and Sahoo 2015). All things considered, the expression patterns of *bHLH* genes in anatomical parts in this study showed that seed parts, having role in emergence and germination, and adult leaves were the main anatomical parts in which *bHLH* genes may play important roles.

The expression profiles of *bHLH* genes investigated in this study showed similar pattern to anatomical parts and developmental stages in terms of perturbation studies (Fig. 8). Depending on perturbation studies in Genevestigator, *bHLH101* and *bHLH39* were clearly upregulated by numerous stimuli and particularly by iron deficiency. Sivitz et al. (2012) reported that *IRT1* and *FRO2*, *MYB72*, and *At3g07720* were found to be targets of *FIT* and *bHLH100/101* mutants showed chlorosis under iron deficiency. Also, they investigated iron regulated genes such as *bHLH038, bHLH039, ZIF1* and *MTP3* in wild-type and double mutant plants. These genes were not reported to be deregulated in the double mutant. Therefore, they concluded that these genes act in iron uptake mechanism *FIT* independent way. *bHLH39* and *bHLH101* were reported to be upregulated in exchange for iron deficiency. Dinneny et al. (2008) found that after 24 h low iron conditions triggered different levels of a high number of gene expressions involving in iron metabolism similar to the salt stress in *Arabidopsis*; however, these genes were mostly

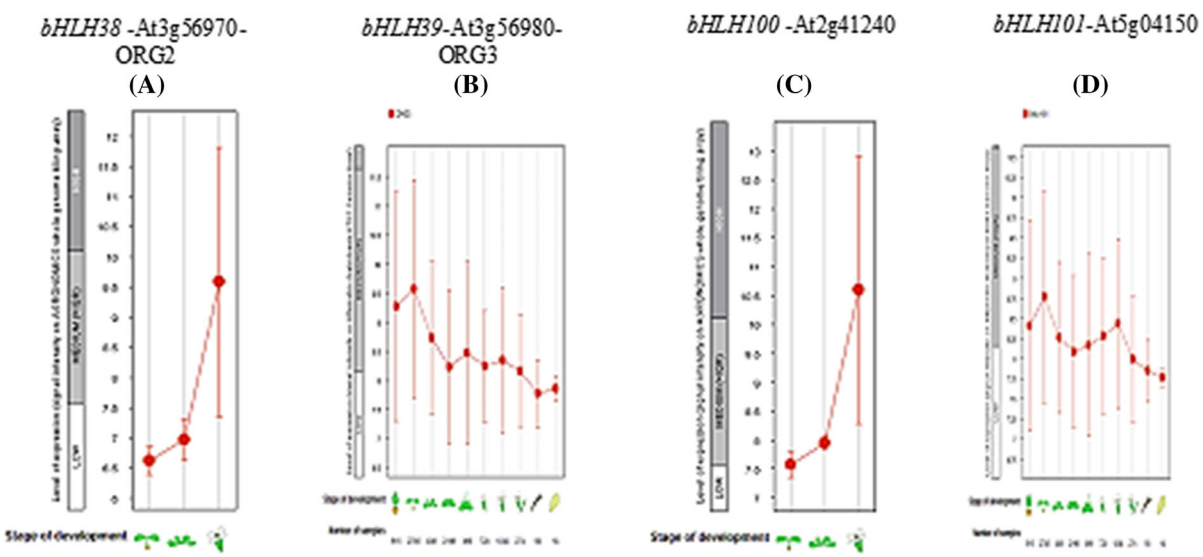

Fig. 6 Gene expression levels of *bHLH* genes in different developmental stages of *Arabidopsis*, including germinated seed, seedling, young rosette, developed rosette, bolting, young flower, developed flower, flower and siliques, mature siliques, and senescence

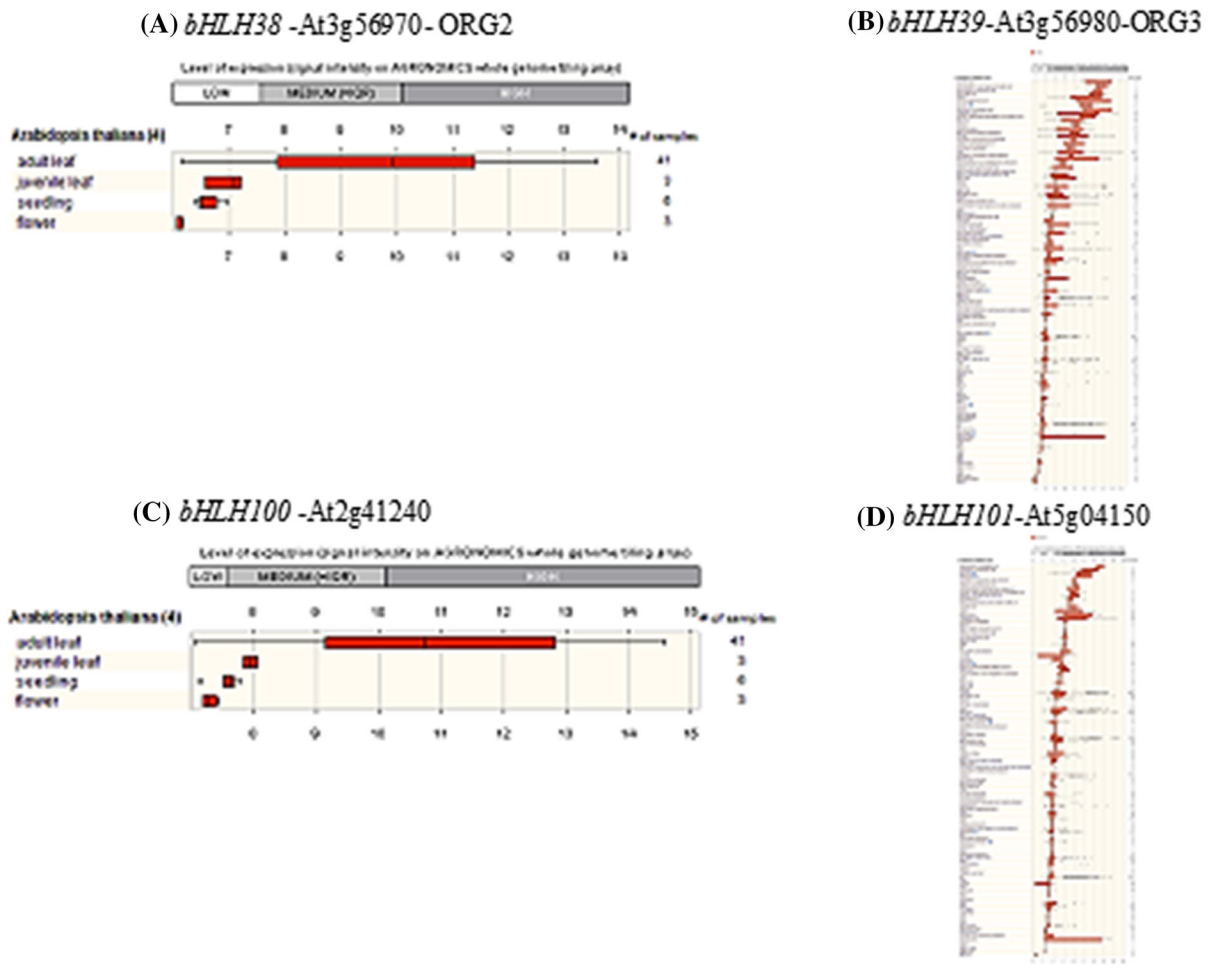

Fig. 7 The expression levels of *bHLH38/39/100/101* genes in different anatomical parts of *Arabidopsis*

upregulated in seeds where iron stored and distributed in matured plants. Buckhout et al. (2009) found that 65 and 79 genes were expressed in roots following 6 and 24 h of removal of iron from hydroponic system and 65% of these genes were identified as the same genes such as *bHLH101* and *bHLH39*. Long et al. (2010) identified *PYE* transcription factor along with *BTS* gene in iron uptake mechanism in *Arabidopsis*. They suggested that *bHLH39* and *bHLH101* were transcribed after *Arabidopsis* exposed to 24 h of iron deficient medium. Stein and Waters (2012) reported that *Arabidopsis* genotypes showed different gene expression profiles under iron deficient conditions, suggesting that genes having role in iron deficiency may have secondary or new roles in regulating iron homeostasis and acquisition of iron. The exposure of these ecotypes to iron deficiency for 24 h increased expression levels of Fe-deficiency-regulated genes such as *bHLH39* and *bHLH101*. Iyer-Pascuzzi et al. (2011) reported sulphur deficient media specifically increased *bHLH39* expression levels. To better understand the mechanism of iron uptake under iron deficient conditions, it is necessary to reveal the relationship between these perturbation experiments and iron uptake as well. In this regard, hormones and some small molecules whose responses were led by environmental conditions were studied by many authors in last years to shed light on iron acquisition mechanism. As a result of these efforts, ethylene, auxin, brassinosteroids, jasmonic acid, ABA, and cytokinins were reported having important roles in iron homeostasis. Lastly, gibberellic acid was suggested to modulate synthesis of BHLH038, BHLH039, FRO2, and IRT1 proteins (Brumbarova

(A) *bHLH38* -At3g56970- ORG2

(B) *bHLH39*-At3g56980-ORG3

(C) *bHLH100* -At2g41240

(D) *bHLH101*-At5g04150

◀**Fig. 8** The expression levels of *Arabidopsis bHLH38/39/100/101* genes in response to different perturbations

et al. 2014). Also, particularly involvement of *bHLH38* in iron uptake mechanism was found subtler. This may be emanated from post-transcriptional changes of bHLH38 protein (Ramamurthy and Waters 2017). In summary, these results suggest that four Ib subgroup *bHLH* genes were co-regulated many genes involved in more than one pathway.

Conclusion

In this study, four Ib subgroup *bHLH* genes into iron acquisition were investigated in A*rabidopsis*, soybean, tomato, rice and maize. Our analyses showed that these genes were not clearly separated, particularly *bHLH101* were evolutionary more distant to other *bHLH* genes. Although *bHLH38/39/100/101* genes were identified as functionally redundant in iron uptake, their secondary and tertiary structures were not similar to each other. Also, *bHLH39* and *bHLH101* may be more acted in iron uptake under iron deficient conditions. *In Arabidopsis,* these genes were upregulated in seedling stage. Also, seed parts and adult leaves were anatomical parts in which these two genes were highly transcribed. Finally, it can be proposed that *bHLH38/39/100/101* genes play important roles in regulating iron homeostasis in studied plants. In addition, these genes can be used as efficient tools for biofortification studies in crops, particularly enriched iron contents.

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaquchi-Shinozaki K (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63–78

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. Plant Cell Physiol 57:e5

Brumbarova T, Bauer P, Ivanov R (2014) Molecular mechanisms governing *Arabidopsis* iron uptake. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2014.11.004

Buckhout TJ, Yang TJ, Schmidt W (2009) Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. BMC Genom 10:147

Celma JR, Pan C, Li W, Lan P, Buckhout TJ, Schmidt W. (2012) The transcriptional response of *Arabidopsis* leaves to Fe deficiency. Front Plant Sci 4:1–10. Article 276

Colangelo EP, Guennoti ML (2004) The essential basic helix-loop-helix protein fit1 is required for the iron deficiency response. Plant Cell 16:3400–34121

Conorton JM, Balk J, Celma JR (2017) Iron homeostasis in plants—a brief overview. Metallomics 9:813

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN (2008) Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. Science 320(5878):942–945

Eisenhaber F. (2006) Prediction of protein function. In: Discovering biomolecular mechanisms with computational biology. Molecular Biology Intelligence Unit. Springer, Boston

Filiz E, Vatansever R, Ozyigit II (2017) Dissecting a co-expression network of basic helix-loop-helix (bHLH) genes from phosphate (Pi)-starved soybean (Glycine max). Plant Gene 9:19–25

Garcia CMH, Finer JJ (2014) Identification and validation of promoters and cis-acting regulatory elements. Plant Sci 217–218:109–119

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD et al (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana, Louisville, pp 571–607

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:1178–1186

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

He H, Yan J, Yu X, Liang Y, Fang L, Scheller HV, Zhang A (2017) The NADPH-oxidase AtRbohI plays a positive role in drought-stress response in *Arabidopsis thaliana*. Biochem Biophys Res Commun 491(3):834–839. https://doi.org/10.1016/j.bbrc.2017.05.131

Heim MA, Jacoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol 20:735–747

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinform. https://doi.org/10.1155/2008/420747

Hudson KA, Hudson ME (2015) A classification of basic helix-loop-helix transcription factors of soybean. Int J Genom. https://doi.org/10.1155/2015/603182

Iyer-Pascuzzi AS, Jackson T, Cui H, Petricka JJ, Busch W, Tsukagoshi H, Benfey PN (2011) Cell identity regulators link development and stress responses in the Arabidopsis root. Dev Cell 21(4):770–782

Jones S (2004) An overview of the basic helix-loop-helix proteins. Genome Biol 5:226

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre^2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858

Kim SA, Guerinot ML (2007) Mining iron: iron uptake and transport in plants. FEBS Lett 581:2273–2280

Kim JK, Cho Y, Lee M, Laskowski RA, Ryu SE, Sugihara K, Kim DS (2015) BetaCavityWeb: a webserver for molecular voids and channels. Nucleic Acids Res 43(W1):W413–W418. https://doi.org/10.1093/nar/gkv360

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2013) CCAAT-box binding transcription factors in plants: y so many? Trends Plant Sci 18(3):157–166

Lescot M, De´hais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30(1):325–327

Li J, Liu B, Cheng F, Wang X, Aarts MGM, Wu J (2014) Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in Brassica Rapa. New Phytol 203:182–194

Liang G, Zhang H, Li X, Ai Q, Yu D (2017) bHLH transcription factor bHLH115 regulates iron homeostasis in *Arabidopsis thaliana*. J Exp Bot 68(7):1743–1755

Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE, Benfey PN (2010) The bHLH transcription factor POPEYE regulates response to iron deficiency in *Arabidopsis* roots. Plant Cell 22:2219–2236

Marschner H (1995) Mineral nutrition of higher plants. Academic Press, London

Murre C, Bain G, Vandijk MA, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR, Stuiver MH (1994) Structure and function of helix-loop-helix proteins. Biochim Biophys Acta 1218:129–135

Nguyen MN, Tan KP, Madhusudhan MS (2011) CLICK - topology-independent comparison of biomolecular 3D structures. Nucleic Acids Res 39:W24–W28

Niu X, Guan Y, Chen S, Li H (2017) Genome-wide analysis of basic helix-loophelix (bHLH) transcription factors in *Brachypodium distachyon*. BMC Genom 18:619

Pires N, Dolan L (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol 27(4):862–874

Ramamurthy RK, Waters BM (2017) Mapping and characterization of the fefe gene that controls iron uptake in melon (*Cucumis melo L.*). Front Plant Sci 8:1003

Rasheed S, Bashir K, Matsui A, Tanaka M, Seki M (2016) Transcriptomic analysis of soil-grown Arabidopsis thaliana roots and shoots in response to a drought stress. Front. Plant Sci. 7:180. https://doi.org/10.3389/fpls.2016.00180

Rizhsky L, Liang H, Mittler R (2003) The water-water cycle is essential for chloroplast protection in the absence of stress. J Biol Chem 278(40):38921–38925

Rout GR, Sahoo S (2015) Role of iron in plant growth and metabolism. Rev Agric Sci 3(1–24):2015

Sivitz AB, Hermand V, Curie C, Vert G (2012) *Arabidopsis* bHLH100 and bHLH101 control iron homeostasis via a FIT-Independent Pathway. PLoS ONE 7(9):e44843

Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28:405–420

Stein RJ, Waters BM (2012) Use of natural variation reveals core genes in the transcriptome of iron-deficient *Arabidopsis thaliana* roots. J Exp Bot 63(2):1039–1055

Sun H, Fan HJ, Ling HQ (2015) Genome-wide identification and characterization of the bHLH gene family in tomato. BMC Genom 16:9

Thomine S, Vert G (2013) Iron transport in plants: better be safe than sorry. Curr Opin Plant Biol 16:322–327

Timothy L, Mikael Bode´n B, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:202–208

Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15:1749–1770

Vatansever R, Filiz E, Ozyigit II (2015) Genome-wide analysis of iron-regulated transporter 1 (irt1) genes in plants. hortic. Environ Biotechnol 56:516–523

Wang H, Klatte M, Jakoby M, Bäumlein H, Weisshaar B, Bauer P (2007) Iron defciency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis thaliana*. Planta 226:897–908

Wang N, Cui Y, Liu Y, Fan H, Du J, Huang Z, Yuan Y, Wu H, Ling HQ (2013) Requirement and functional redundancy of Ib Subgroup bHLH proteins for iron deficiency responses and uptake in *Arabidopsis thaliana*. Mol Plant 6(2):503–513

Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 31(13):3316–3319

Wong DCJ, Gutierrez RL, Gambetta GA, Castellarin SD (2017) Genome-wide analysis of cis-regulatory element structure and discovery of motif-driven gene co-expression networks in grapevine. DNA Res 24(3):311–326

Xiang Q (2015) Molecular genetic aspects of iron and copper cross-talk in *Arabidopsis thaliana*. Dissertation, University of Nebraska-Lincoln

Xing J, Wang T, Ni Z (2015) Epigenetic regulation of iron homeostasis in *Arabidopsis*. Plant Signal Behav 10:12

Yousefi M, Nematzadeh GA, Askari H, Nasiri N (2012) Bioinformatics analysis of promoters cis elements and study on genes co-expression of oxidative defense pathway in *Oryza sativa* L. Plant. Int J Agric Crop Sci 4(17):1240–1245

Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64:643–651

Zhang J, Liu B, Mengshu L, Feng D, Jin H, Wang P, Liu J, Xiong F, Wang J, Wang HB (2015a) The bHLH transcription factor bHLH104 interacts with IAA-leucine resistant3 and modulates iron homeostasis in *Arabidopsis*. Plant Cell 27:787–805

Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, Chen S (2015b) Genome-wide characterization and analysis of *bHLH* transcription factors related to tanshinone biosynthesis in Salvia miltiorrhiza. Sci Rep 5:1124